1. A.

Text

Description automatically generated

B.

A picture containing chart

Description automatically generated

The biggest anomaly I see is passenger id. It does not translate well into histograms. The passenger fare could also be trained smaller so there is a more even distribution.

C. Those features are not important for understanding the data in this instance. Home.dest could be important if you wanted to get a distribution of where everyone was going afterward, but the boat was going to the same place with everyone on it so most destinations would be the same. They are also all on the same boat and in the same body of water. These would be redundant to keep in the data set.

1. A. I believe a feature that would be useful for survival is the location on the boat in comparison to access to life boats. The argument would be that if a passenger is closer on the ship to a life boat, they are more likely to be granted access and survive. This could be achieved using cabin and class in the data set, but would also need to have a distance from life boat added.

B.

![Graphical user interface

Description automatically generated]()

C. The two most important features are age and survival because they show age variations. With these age variations, we can discern for future reference who would survive a similar situation. It seems that younger individuals were more likely to survive, with some outliers. By using these features, we can also get an accurate image of the data. They are the most stable features.

3. A.

![Table

Description automatically generated]()

1. B.

![Text, letter

Description automatically generated]()

I received this error running the code at the bottom of the starter.

![Graphical user interface, text, application, email

Description automatically generated]()

3.C.

The training data seemed to be more accurate than the test data. That being said, both test and training sets had relatively good accuracy. It seems that this is relatively stable.